

СЕМАНТИКО-КОГНИТИВНЫЙ МЕТОД ВЫЯВЛЕНИЯ ЗНАНИЙ ГЛОССАРИЯ

Петрасова С. В.

Научный руководитель – д.т.н., доц. Хайрова Н. Ф.

Национальный технический университет

"Харьковский политехнический институт"

(61002, Харьков, ул. Пушкинская, 79/2 каф. Интеллектуальных
компьютерных систем, тел. (057) 707-63-60)

e-mail: svetapetrasova@gmail.com

The given work is devoted to the method of the automatic construction of the domain semantic network on the basis of glossary knowledge as the most concentrated natural language text. In this work, the cognitive paradigm of the identification of implicit relations' category of equivalence using factorization of concepts' domain has been considered. Patterns of lexical sets have been used for the formalization of the explicit categories of concepts' relations of the semantic network such as belonging to the class, hypernymy, hyponymy and myronymy.

Современный уровень разработок по созданию хранилищ знаний и их систематизации требует внедрения универсальных интеллектуальных систем, основным компонентом которых является модель экстракции и представления знаний.

Одной из наиболее известных форм представления знаний является аппарат семантических сетей. Основная идея такого представления состоит в описании семантики отражаемой ПО с помощью сетевых структур, основанных на теории графов.

В данном исследовании предлагается семантико-когнитивный метод автоматического выявления отношений объектов семантической сети, основанный на знаниях глоссариев, выражаемых дефинициями терминов данных объектов [1].

Для построения логической схемы выделения семантически близких терминов вводится метрическое пространство лингвистических смысловых единиц Θ , определяемое как множество лингвистических единиц лексикона T , на котором грамматические правила задают отношения между единицами, выступающими ограничениями для корректных синтаксических структур [2]. Мера семантической близости $f(t', t'')$ между двумя лингвистическими единицами t' и t'' определяется формулой:

$$f(t', t'') = \frac{2|d_1 \cap d_2|}{|d_1| + |d_2|},$$

где $d_1 \cap d_2$ – общие термины определений, а $|d_1| + |d_2|$ – все термины определений d_1 и d_2 .

Будем говорить, что две лингвистические единицы связаны в одном своем семантическом значении, и писать $(t_i, d_i) \sim (t_j, d_j)$, если только $F(t_i, d_i) = F(t_j, d_j)$. Легко показать, что отношение \sim выражает эквивалентность и факторизует пространства лингвистических смысловых единиц Θ и исследуемых связных текстов Ω , разбивая их на классы эквивалентности [3].

Отношение эквивалентности позволяет организовать различные лингвистические единицы в классы эквивалентности, определяющие смысловую близость лингвистических единиц, тем самым, факторизовав пространство концептов, выражаемых знаками лингвистических смысловых единиц.

Смысловая близость лингвистических единиц включает такие категории семантических отношений, как: принадлежность к классу, гиперонимия, гипонимия и меронимия. Для определения описанных типов отношений к лингвистическим единицам одного класса эквивалентности применяются шаблоны лексических последовательностей:

$$NN_1 \rightarrow \text{Rel } l_z \rightarrow NN_2,$$

где NN_1 и NN_2 – связанные концепты, представленные ключевыми словами и словосочетаниями глоссариев, $\text{Rel } l$ – лексические цепочки, выражающие отношения z :

z =отношение классификации, $\text{Rel}=\{ \text{“вважається”} \}$.

z =отношение гиперонимии, $\text{Rel}=\{ \text{“сукупність”, “комплекс”, “набір”, “сімейство”} \}$.

z =отношение гипонимии, $\text{Rel}=\{ \text{“наприклад”, “тип”, “(різно)вид”, “екземпляр”} \}$.

z =отношение меронимии, $\text{Rel}=\{ \text{“частина”, “елемент”} \}$.

В результате исследования предложен метод автоматического построения семантической сети, основанный на использовании глоссариев как естественно-языкового текстов, наиболее полно концентрирующих знания ПО.

Для выявления явных категорий межконцептуальных отношений к лексическим единицам, определенным в единый класс эквивалентности, применяются шаблоны лексических последовательностей.

Литература:

1. Кобозева И. М. Лингвистическая семантика: Учебное пособие. / И. М. Кобозева. – М. : Эдиториал УРСС, 2000. – 352 с.
2. Хайрова Н. Ф. Определение семантической близости на основе когнитивного подхода. / Н. Ф. Хайрова, Н. В. Шаронова, Н. В. Борисова // Бионика интеллекта: науч.-техн. журнал, 2013.
3. Бондаренко М. Ф. Теория интеллекта: учебник / М. Ф. Бондаренко, Ю. П. Шабанов-Кушнаренко – Харьков: Комп. СМІТ, 2007. – 576 с.